Locally Equivalent Weights for Bayesian MrP

Ryan Giordano, Alice Cima, Erin Hartman, Jared Murray, Avi Feller

UT Austin Statistics Seminar
September 2025

What are we weighting for?[1]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

What are we weighting for?[1]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{N_T}\sum_{j=1}^{N_T} \mathbf{x}_j = \frac{1}{N_S}\sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Such weights satisfy "covariate balance" for **x**.

You can check covariate balance for any calibration weighting estimator, and any function $f(\mathbf{x})$.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

What are we weighting for?[1]

$$\text{Target average response} = \frac{1}{N_T}\sum_{j=1}^{N_T} y_j \approx \frac{1}{N_S}\sum_{i=1}^{N_S} w_i y_i = \text{Weighted survey average response}$$

We can't check this, because we don't observe y_j. But we can check whether:

$$\frac{1}{N_T}\sum_{j=1}^{N_T} \mathbf{x}_j = \frac{1}{N_S}\sum_{i=1}^{N_S} w_i \mathbf{x}_i$$

Such weights satisfy "covariate balance" for **x**.

You can check covariate balance for any calibration weighting estimator, and any function $f(\mathbf{x})$.

Even more, covariate balance is the criterion for a popular class of calibration weight estimators:

Raking calibration weights

"Raking" selects weights that

- Are as "close as possible" to some reference weights
- Under the constraint that they balance some selected regressors.

[1] Pun attributable to Solon, Haider, and Wooldridge (2015)

Balance checks as sensitivity analysis

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance checks as sensitivity analysis

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (informal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

Balance checks as sensitivity analysis

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{N_T}\sum_{j=1}^{N_T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{y}, y$:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} \overset{\text{check}}{\approx} \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

Balance checks as sensitivity analysis

One reason to balance $f(\mathbf{x})$ is because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

Balance–informed sensitivity check (BISC) (formal)

Pick a small $\delta > 0$ and an $f(\cdot)$. Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{N_T}\sum_{j=1}^{N_T}\left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j)$$

Then, check whether your estimator $\hat{\mu}(\cdot)$ produces the same change for observed $\tilde{y}, y$:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} \overset{\textbf{check}}{\approx} \delta\frac{1}{N_T}\sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

When $\hat{\mu}(\cdot) = \hat{\mu}_{\mathrm{CW}}(\cdot)$, BISC recovers the standard covariate balance check.

We will use $\hat{\mu}(\cdot) = \hat{\mu}_{\mathrm{MrP}}(\cdot)$.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y)$.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$.

Problem: $\hat{\mu}_{\mathrm{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$ may be small.

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$.

Problem: $\hat{\mu}_{\mathrm{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the first-order Taylor series approximation

$$\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y) \approx \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}}(\tilde{y}_i - y_i) \quad \text{where} \quad w_i^{\mathrm{MrP}} := \frac{d}{dy_i}\hat{\mu}_{\mathrm{MrP}}(y).$$

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$.

Problem: $\hat{\mu}_{\mathrm{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the first-order Taylor series approximation

$$\hat{\mu}_{\mathrm{MrP}}(\tilde{y}) - \hat{\mu}_{\mathrm{MrP}}(y) \approx \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}}(\tilde{y}_i - y_i) \quad \text{where} \quad w_i^{\mathrm{MrP}} := \frac{d}{dy_i}\hat{\mu}_{\mathrm{MrP}}(y).$$

Computation: The weights are given by weighted averages of posterior covariances[2].

They can be easily computed with standard software[3] **without re–running MCMC**.

[2] **G.**, Broderick, and Jordan 2018.
[3] We use `brms` (Bürkner 2017).

BISC for MrP

Suppose I have $\tilde{y}$ such that $\mathbb{E}[\tilde{y}|\mathbf{x}] = \mathbb{E}[y|\mathbf{x}] + \delta f(\mathbf{x})$.
Now I need to evaluate $\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y)$.

Problem: $\hat{\mu}_{\text{MrP}}(\cdot)$ is computed with MCMC.

- Each MCMC run typically takes hours, and
- Output is noisy, and $\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y)$ may be small.

MrP Local Equivalent Weights (MrPlew)

Form the first-order Taylor series approximation

$$\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y) \approx \sum_{i=1}^{N_S} w_i^{\text{MrP}} (\tilde{y}_i - y_i) \quad \text{where} \quad w_i^{\text{MrP}} := \frac{d}{dy_i} \hat{\mu}_{\text{MrP}}(y).$$

Use in BISC: For a wide set of judiciously chosen $f(\cdot)$, check

$$\delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \overset{\text{check}}{\approx} \delta \frac{1}{N_T} \sum_{j=1}^{N_T} f(\mathbf{x}_j).$$

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** Two approaches:

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}\left[y|\mathbf{x}\right]$
   - E.g. Posterior mean, or
   - Shrunken posterior mean, or
   - Some values that gives the same posterior
2. Take $u_n \overset{iid}{\sim} \text{Unif}(0,1)$
3. Assume $y_n = \mathbb{I}\left(u_n \le \hat{m}(\mathbf{x}_n)\right)$
4. Draw $u_n | y_n$
5. Set $\tilde{y}_n = \mathbb{I}\left(u_n \le \hat{m}(\mathbf{x}_n) + \delta \mathbf{x}_n\right)$

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

Generating $\tilde{y}$

- We have defined BISC in terms of $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$
- We have approximated $\hat{\mu}_{\text{MrP}}(\tilde{y}) - \hat{\mu}_{\text{MrP}}(y)$ for $\tilde{y} \approx y$

How to get such a $\tilde{y}$? **Recall y is binary!** Two approaches:

Option 1: Force $\tilde{y}$ to be binary.

1. Make *some* guess $\hat{m}(\mathbf{x}) \approx \mathbb{E}\left[y|\mathbf{x}\right]$
   - E.g. Posterior mean, or
   - Shrunken posterior mean, or
   - Some values that gives the same posterior
2. Take $u_n \overset{iid}{\sim} \text{Unif}(0,1)$
3. Assume $y_n = \mathbb{I}\left(u_n \leq \hat{m}(\mathbf{x}_n)\right)$
4. Draw $u_n | y_n$
5. Set $\tilde{y}_n = \mathbb{I}\left(u_n \leq \hat{m}(\mathbf{x}_n) + \delta \mathbf{x}_n\right)$

- Realistic
- Have to pick $\hat{m}(\mathbf{x})$
- $\tilde{Y} - Y_{\mathcal{S}}$ not infinitesimally small
- **Sanity check for theory**

Option 2: Allow $\tilde{y}$ to take generic values.

1. Set $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

- Not realistic
- No additional assumptions
- $\tilde{Y} - Y_{\mathcal{S}}$ may be infinitesimally small
- **Use for theory**

BISC Theorem: (sketch)

Take $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}_{\mathrm{MrP}}(Y_{\mathcal{S}}) - \hat{\mu}_{\mathrm{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \right| = \text{Small?}$$

[2] Donsker with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[3] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

Theory

BISC Theorem: (sketch)

Take $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}_{\text{MrP}}(Y_{\mathcal{S}}) - \hat{\mu}_{\text{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2)$$

[2] Donsker with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[3] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

Theory

BISC Theorem: (sketch)

Take $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\left| \hat{\mu}_{\mathrm{MrP}}(Y_{\mathcal{S}}) - \hat{\mu}_{\mathrm{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\mathrm{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

[2] Donsker with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[3] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

BISC Theorem: (sketch)

Take $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}_{\text{MrP}}(\tilde{Y}_{\mathcal{S}}) - \hat{\mu}_{\text{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

For a very broad class[2] of $\mathcal{F}$.

Uniformity justifies searching for “imabalanced” f.

[2] Donsker with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[3] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

BISC Theorem: (sketch)

Take $\tilde{y}_n = y_n + \delta f(\mathbf{x}_n)$.

We state conditions for Bayesian hierarchical logistic regression under which

$$\sup_{f \in \mathcal{F}} \left| \hat{\mu}_{\text{MrP}}(Y_{\mathcal{S}}) - \hat{\mu}_{\text{MrP}}(Y_{\mathcal{S}}) - \delta \sum_{i=1}^{N_S} w_i^{\text{MrP}} f(\mathbf{x}_i) \right| = O(\delta^2) \text{ as } N \to \infty$$

For a very broad class[2] of $\mathcal{F}$.

Uniformity justifies searching for "imabalanced" f.

The uniformity result builds on our earlier work on uniform and finite–sample error bounds for Bernstein–von Mises theorem–like results[3].

[2] Donsker with uniformly bounded $\mathbb{E}\left[\mathbf{x} f(\mathbf{x})\right]$.

[3] **G.** and Broderick 2024; Kasprzak, **G.**, and Broderick 2025.

Real Data: Marital Name Change Survey

Analysis of changing names after marriage[4].

- **Target population:** ACS survey of US population 2017–2022[5]
- **Survey population:** Marital Name Change Survey (from Twitter)[6]
- **Respose:** Did the female partner keep their name after marriage?
- For regressors, use bins of age, education, state, and decade married.

Survey observations:	$N_S = 4,364$
Target observations (rows):	$N_T = 4,085,282$

Uncorrected survey mean:	$\frac{1}{N_S}\sum_{i=1}^{N_S} y_i = 0.462$
Raking:	$\hat{\mu}_{\text{CW}} = 0.263$
MrP:	$\hat{\mu}_{\text{MrP}} = 0.288 \quad (\text{Post. sd} = 0.0169)$

[4] Based on Alexander (2019).

[5] Ruggles et al. 2024.

[6] Cohen 2019.

Covariate balance for primary effects

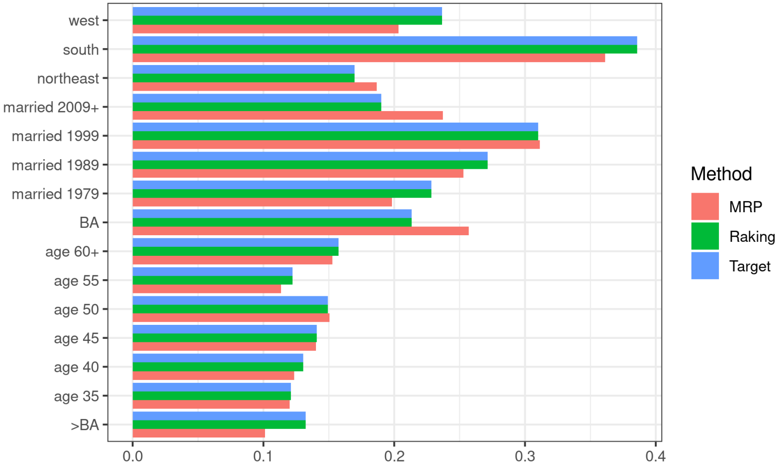

Figure 1: Imbalance plot for primary effects

Covariate balance for interaction effects

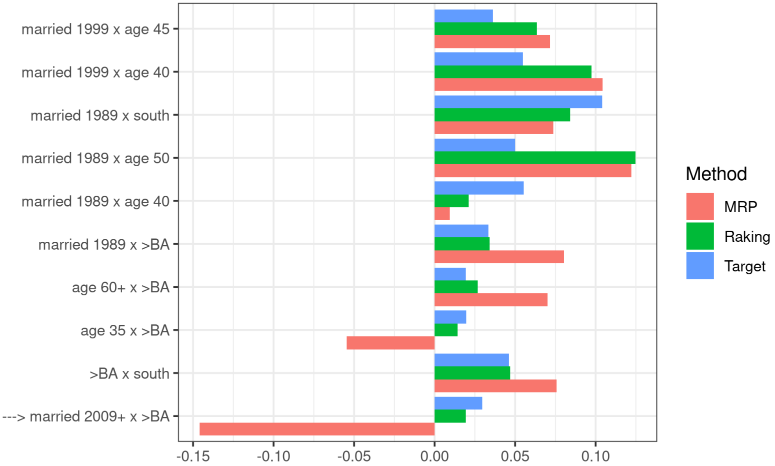

Figure 2: Imbalance plot for select interaction effects

Predictions

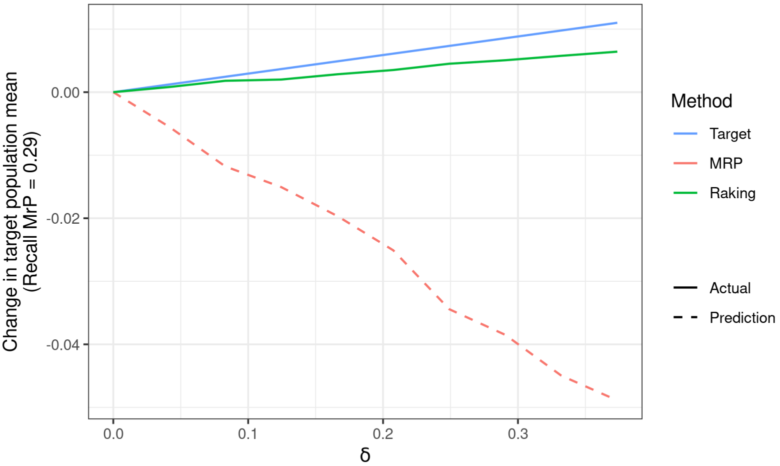

Figure 3: Predictions for the name change dataset

Predictions and actual MCMC results

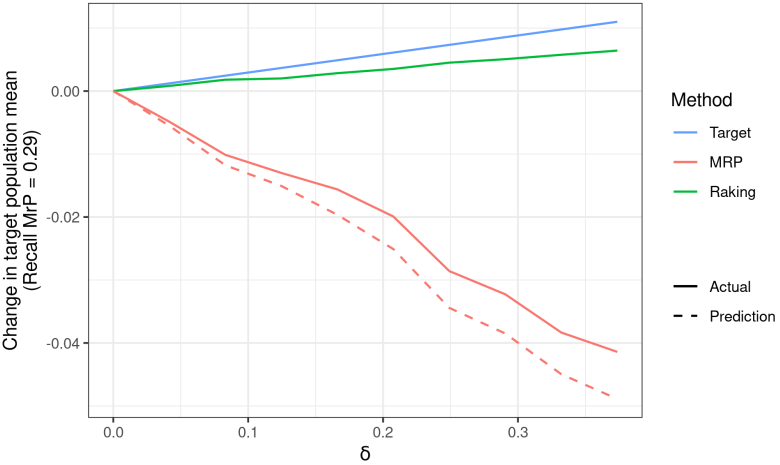

Figure 4: Predictions and refit for the name change dataset

Running ten MCMC refits: 10 hours Computing approximate weights: 16 seconds

Real Data: Lax Philips

Analysis of national support for gay marriage.[7]

- **Target population:** US Census Public Use Microdata Sample 2000
- **Survey population:** Combined national-level polls from 2004
- **Respose:** "Do you favor allowing gay and lesbian couples to marry legally?"
- For regressors, use race, gender, age, education, state, region, and continuous statewide religion and political characteristics, including some analyst–selected interactions.

Survey observations:	$N_S = 6,341$
Target observations (rows):	$N_T = 9,694,541$

Uncorrected survey mean:	$\frac{1}{N_S}\sum_{i=1}^{N_S} y_i = 0.333$
Raking:	$\hat{\mu}_{\text{CW}} = 0.33$
MrP:	$\hat{\mu}_{\text{MrP}} = 0.337 \quad (\text{Post. sd} = 0.039)$

[7] Based on Kastellec, Lax, and Phillips (2010), see also Lax and Phillips (2009).

Covariate balance for primary effects

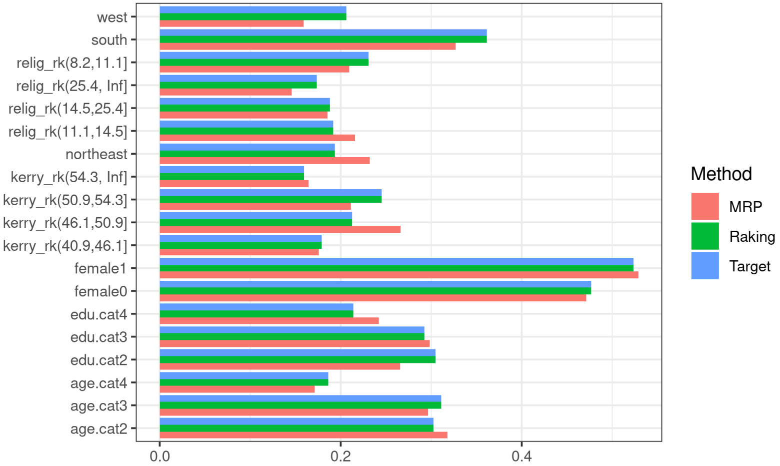

Figure 5: Imbalance plot for primary effects

Covariate balance for interaction effects

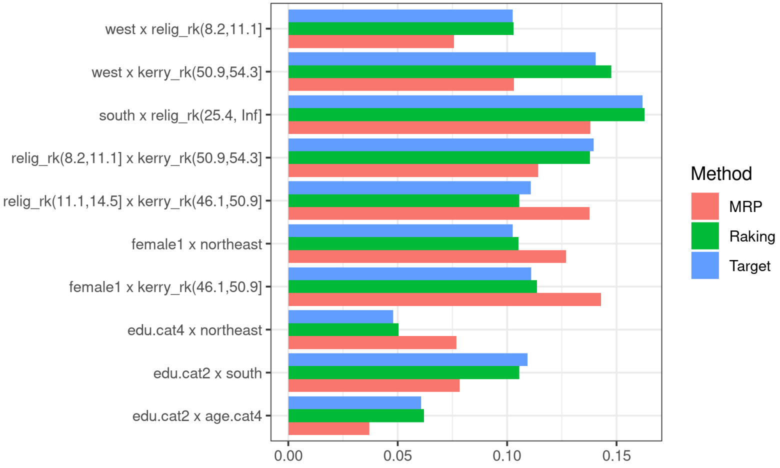

Figure 6: Imbalance plot for select interaction effects

Predictions

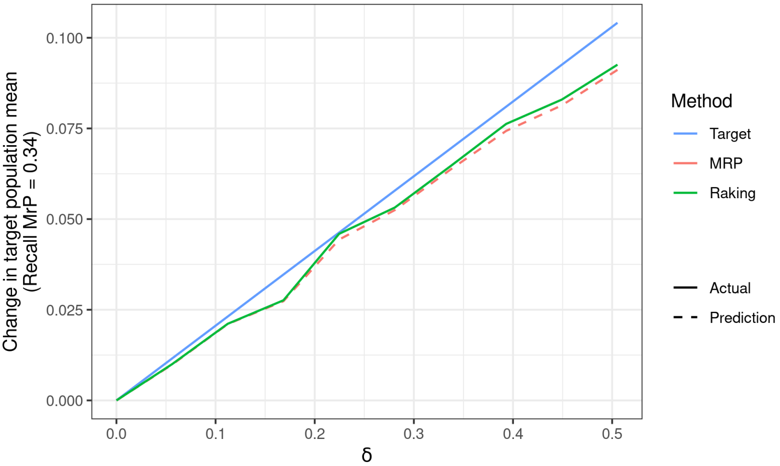

Figure 7: Predictions for the gay marriage dataset

Predictions and actual MCMC results

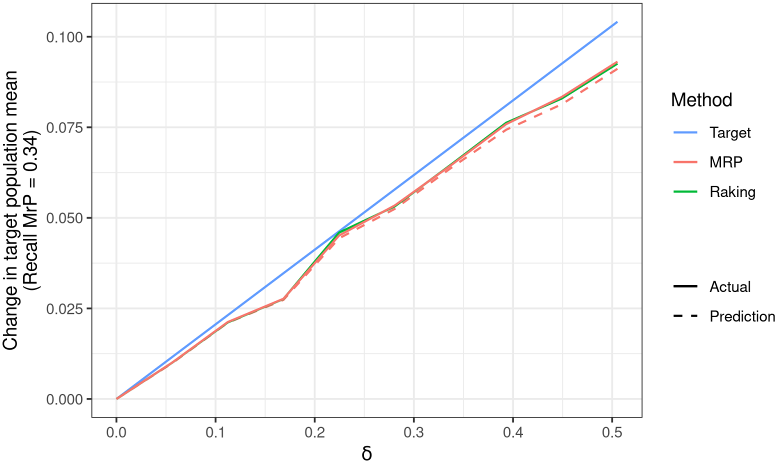

Figure 8: Predictions and refit for the gay marriage dataset

Running ten MCMC refits: 11 hours Computing approximate weights: 23 seconds

Predictions

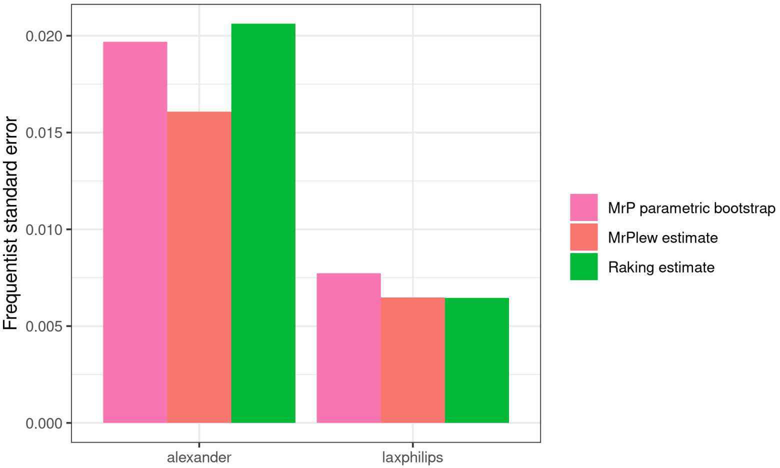

Figure 9: Frequentist standard deviation estimates

References

Alexander, M. (2019). *Analyzing name changes after marriage using a non-representative survey*. URL: https://www.monicaalexander.com/posts/2019-08-07-mrp/.

Bürkner, Paul-Christian (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan". In: *Journal of Statistical Software* 80.1, pp. 1–28. DOI: 10.18637/jss.v080.i01.

Cohen, P. (Apr. 2019). *Marital Name Change Survey*. DOI: 10.17605/OSF.IO/UZQDN. URL: osf.io/uzqdn.

G. and T. Broderick (2024). *The Bayesian Infinitesimal Jackknife for Variance*. arXiv: 2305.06466 [stat.ME]. URL: https://arxiv.org/abs/2305.06466.

G., T. Broderick, and M. I. Jordan (2018). "Covariances, robustness and variational bayes". In: *Journal of machine learning research* 19.51.

Kasprzak, M., **G.**, and T. Broderick (2025). *How good is your Laplace approximation of the Bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences*. arXiv: 2209.14992 [math.ST]. URL: https://arxiv.org/abs/2209.14992.

Kastellec, J., J. Lax, and J. Phillips (2010). "Estimating state public opinion with multi-level regression and poststratification using R". In: *Unpublished manuscript, Princeton University* 29.3.

Lax, J. and J. Phillips (2009). "Gay rights in the states: Public opinion and policy responsiveness". In: *American Political Science Review* 103.3, pp. 367–386.

Ruggles, S. et al. (2024). *IPUMS USA: Version 15.0 [dataset]*. DOI: 10.18128/D010.V15.0. URL: https://usa.ipums.org.

Solon, G., S. Haider, and J. Wooldridge (2015). "What are we weighting for?" In: *Journal of Human resources* 50.2, pp. 301–316.